

Post-Capture Lighting Manipulation using Flash Photography

ZHUO HUI, Carnegie Mellon University

KALYAN SUNKAVALLI, Adobe Research

SUNIL HADAP, Adobe Research

ASWIN C SANKARANARAYANAN, Carnegie Mellon University

Real-world lighting often consists of multiple illuminants with different colors. Separating and manipulating these illuminants in post-process is a challenging problem that requires either significant manual input or calibrated scene geometry and lighting. In this work, we leverage a flash/no-flash image pair to analyze and edit scene illuminants based on their color differences. We derive a novel physics-based relationship between color variations in the observed flash/no-flash intensities and the chromaticities and surface shading corresponding to individual scene illuminants. Our technique uses this constraint to automatically separate an image into constituent images lit by each illuminant. Each light component can then be edited independently to enable applications like white-balance, lighting editing, and intrinsic image decompositions. We demonstrate that this technique outperforms state-of-the-art techniques for some applications and enables other applications that are not possible with previous work.

CCS Concepts: • **Computing methodologies** → **Image manipulation**; *Computational photography*;

ACM Reference format:

Zhuo Hui, Kalyan Sunkavalli, Sunil Hadap, and Aswin C Sankaranarayanan. 2017. Post-Capture Lighting Manipulation using Flash Photography. *ACM Trans. Graph.* 1, 1, Article 1 (April 2017), 13 pages.
DOI: 00000001.0000001_2

1 INTRODUCTION

Advances in computational photography have made it possible to edit aspects of the capture process, like the focal plane and aperture [32] in post-process. However, post-process control of scene illumination has stubbornly remained out of the reach of the average consumer. Real-world lighting is often a mixture of multiple illuminants, and the contribution of each illuminant to a scene point is a complex function of scene and lighting geometry. This can be seen in Figure 1(a), where every pixel in this scene is lit by a spatially-varying mixture of cool outdoor illumination and warm indoor lighting. Recovering these contributions is a severely ill-posed problem that has previously been solved using extensive manual annotation [9–11] or access to calibrated scene and lighting information [16, 17]. However, different real-world illuminants often have distinct chromaticities. For example, outdoor illumination – both sunlight and skylight – differ in color temperature from indoor illuminants like tungsten and incandescent lights. Illuminants with different colors induce different color variations in observed intensities; however these color variations are conflated with surface albedo variations, making lighting inference challenging.

In this paper, we take a step towards post-capture control of the illumination in a Lambertian scene using flash photography, i.e.,

two photographs acquired with and without the use of the camera flash. We achieve this by separating the no-flash photograph into multiple photographs, one each for the scene illuminated by lighting of a distinct chromaticity, up to a maximum of three different illuminants (Fig. 1(b)). The key insight behind our technique is that flash photography images the scene under a single illuminant (whose color we can pre-calibrate), thereby enabling us to infer the chromaticity of the surface albedo. Based on this, we derive a novel albedo-invariant, which we call the *hull constraint*, that relates the colors and the relative per-pixel contributions of each scene illuminant to the observed intensities in the no-flash photograph. The Hull constraint is independent of the geometry of the scene and the lighting, and as we show in this paper, generalizes to point and area sources as well as near and distant lighting.

The Hull constraint provides an elegant and robust framework for analyzing scene illumination. In particular, it enables us to *automatically* separate the no-flash photograph into multiple images – each illuminated by the individual light colors. This, in turn, enables a wide-range of automatic, post-capture lighting manipulation applications including image white-balancing under complex mixed illumination, the editing of the color and brightness of the separated illuminants, two-shot intrinsic images, and photometric stereo (see Fig. 1(c,d,e)); we outperform the state-of-the-art in each of the applications, often by significant margins.

Contributions. We propose a flash photography-based algorithm to analyze spatially-varying, mixed illumination without the need for any scene calibration or user input. In particular, we make the following contributions:

1. We introduce a novel algorithm to automatically separate an image into its constituent illuminants using flash photography.
2. We present a theoretical analysis of this technique that explores the robustness and limitations of the algorithm.
3. We leverage these separated images to enable a number of vision and graphics applications including white balance, light editing, intrinsic images, and photometric stereo. In each of these cases, our technique outperforms the state-of-the-art.

Limitations. Our technique relies on differences in color to resolve light sources; hence, two light sources with identical chromaticity will be grouped together as one source with a more complex spatial geometry. We can only separate up to light sources of three distinct chromaticities – a limitation that directly stems from the use of cameras with three-color channels. We assume that the scene is Lambertian and that the flash illuminates all scene points. Finally, we require that the flash/no-flash images are in alignment.



Fig. 1. The room in this example is lit by the warm indoor lights on the top and cool outdoor lighting entering from the left. In this paper, we show that by capturing an additional flash image (a, right), we can automatically separate the no-flash image (a, left) into per-illuminant images (b). This enable a number of applications, including white-balancing the image (c), post-capture light editing to change the mood to the scene (d), and intrinsic image decompositions that allow us to realistically edit textures in the image (e).

2 RELATED WORK

In this section, we review previous work on illumination analysis as well as prior applications of flash photography.

2.1 Lighting analysis

Image formation involves complex interactions between illumination, scene geometry and surface reflectance. As a result, estimating and editing the illumination requires us to invert the image formation process to recover all three factors; this is extremely challenging since all these effects are combined into a single observation making the problem highly ill-posed. Prior work in this area can be broadly categorized into two groups. The first group relies on active illumination to probe the scene and estimate scene properties. The second group relies on priors on illumination (and in some cases, also on geometry and reflectance) to infer scene properties. We discuss these two classes of techniques in this section.

Active illumination. Active illumination methods use controlled illumination to probe and infer scene properties. Controlled capture setups like a light stage [17] are used to capture images of a person or a scene under all incoming light directions. Once captured, these datasets can be used to re-render photorealistic images under arbitrary illumination [18, 36]. However, light stages are not easily available to the vast majority of users. Projector-camera systems have been used extensively to probe and separate light transport in a scene [31]; however, even such systems are sufficiently complex often requiring simultaneous control both the projector and the camera. Similarly, photometric stereo methods are used to capture images under varying illumination that are subsequently used to reconstruct surface normals and reflectance properties [45]. These reconstructions can then be used for relighting applications. However, even these approaches require calibrated capture that is typically not possible in the wild. In contrast, we propose a simple capture process that uses a camera flash to enable a number of illumination analysis, editing and reconstructions tasks including two-shot reflectance estimation and photometric stereo.

Passive illumination. Passive illumination methods aim to estimate scene properties from images captured as-is under natural illumination. Because scene reconstruction from one (or few) image(s) is ill-posed, these techniques often rely on strong priors on

scene geometry, reflectance, and illumination. For example, Barron and Malik [5] demonstrate impressive results for shape, reflectance, and illumination estimation for single object captured under low-frequency distant lighting. Similarly, intrinsic image algorithms [6] that separate images into reflectance and shading components often assume that there is a single (usually white) illuminant and the induced shading is spatially smooth [7, 22, 28, 39, 48]. While these techniques work for some scenarios, real world scenes often have complex geometry and multiple spatially-varying light sources that cannot be model using simple priors. We demonstrate that an additional flash image is sufficient to make this problem well-constrained without relying on strong scene priors and can lead to more accurate estimates of scene properties like shading, reflectance, and shape.

Another class of techniques rely on user input to resolve ambiguities in reflectance and illumination and have been used to compute intrinsic decompositions [9, 10] and white balance under mixed illumination [11, 25]. However, the amount of user interaction required to produce high-quality results in complex scenes is significant. In contrast, our technique only requires an additional image, but is completely automatic and produces state-of-the-art results for applications like white balance and intrinsic image decomposition.

2.2 Flash photography

Flash (or flash/no-flash) photography refers to techniques that capture two images of a scene – with and without flash illumination. Originally proposed to denoise images captured in low-light conditions [19, 37], flash photography has been subsequently used for deblurring [49], artifact removal [19?], non-photorealistic rendering [38], foreground segmentation [41] and matting [42]. More recently Hui et al. [26] demonstrate the use of flash photography for white balancing under mixed illumination. Our approach is similar in spirit to this work in that we leverage the flash/no-flash image pair to obtain an albedo invariant image that is used for lighting analysis. Going beyond that, we derive a novel physics-based constraint that enables explicit separation of the contribution of different light sources at each pixel. Our analysis leads to more accurate white-balance results, and enables a number of applications that are not

possible with prior work, including light editing, intrinsic images, and two-shot photometric stereo.

3 THE HULL CONSTRAINT FOR MIXED ILLUMINATION

Given an image of a scene lit by a mixture of illuminants – the *no-flash* image – our goal is to estimate the contribution of each illuminant to the observed pixel intensities. Doing this from a single image is severely under-constrained since both the surface albedo and incident illumination at every pixel are unknown. In order to address this, we capture an additional *flash* image where the scene is also lit by a flash, whose color we have pre-calibrated. In this section, we set up the image formation model and derive a novel constraint between the observed no-flash/flash pixel intensities and the contributions of each scene illuminant to the scene appearance.

3.1 Problem setup and image formation

We assume that the scene is Lambertian and is imaged by a three-channel color camera. Further, we assume that the scene is lit by N light sources whose chromaticities are denoted as $\{\ell_1, \ell_2, \dots, \ell_N\}$, where $\ell_i = [\ell_i^r, \ell_i^g, \ell_i^b] \in \mathbb{R}^3$ with $\sum_x \ell_i^x = 1$; hence, the light source chromaticities are points on the 2D simplex. We assume that the light source chromaticities are unique, i.e., $i \neq j, \ell_i \neq \ell_j$. Given this setup, the intensity observed at a pixel \mathbf{p} in the no-flash photograph I_{nf} is given by:

$$I_{nf}^c(\mathbf{p}) = \rho^c(\mathbf{p}) \sum_{i=1}^N \lambda_i(\mathbf{p}) \ell_i^c, \quad \text{for } c \in \{r, g, b\}, \quad (1)$$

where $\rho(\mathbf{p}) = [\rho^r(\mathbf{p}), \rho^g(\mathbf{p}), \rho^b(\mathbf{p})]$ is the three-color albedo. The term $\lambda_i(\mathbf{p})$ is the shading observed at pixel \mathbf{p} due to the i -th light source multiplied by the light-source brightness. Recall, that we identify light sources based on their color and hence, two sources with the same color are clustered together; as a consequence, the shading term $\lambda_i(\mathbf{p})$ has a complex dependence on the lighting geometry and does not have a simple analytical form.

Similarly, the intensity observed at pixel \mathbf{p} in the flash photograph I_f is given by:

$$I_f^c(\mathbf{p}) = I_{nf}^c(\mathbf{p}) + \rho^c(\mathbf{p}) \lambda_f(\mathbf{p}) \ell_f^c, \quad \text{for } c \in \{r, g, b\}, \quad (2)$$

where $\lambda_f(\mathbf{p})$ denotes the shading at \mathbf{p} caused by the flash, and the chromaticity of the flash $\ell_f = [\ell_f^r, \ell_f^g, \ell_f^b]$ is assumed to be known via a one-time calibration process.

3.2 The Hull Constraint

Based on the image formation model described above, we now derive the central idea of our work – a novel constraint that encodes both the scene light source chromaticities as well as their contributions to each image pixel.

The flash photograph can be used to estimate the albedo chromaticity, as described in Hui et al. [26]. The pure-flash image I_{pf} is obtained by subtracting the no-flash image from the flash image:

$$I_{pf}^c(\mathbf{p}) = I_f^c(\mathbf{p}) - I_{nf}^c(\mathbf{p}) = \rho^c(\mathbf{p}) \lambda_f(\mathbf{p}) \ell_f^c.$$

Dividing by the pre-calibrated flash chromaticity, ℓ_f , gives us:

$$\alpha^c(\mathbf{p}) = I_{pf}^c(\mathbf{p}) / \ell_f^c = \rho^c(\mathbf{p}) \lambda_f(\mathbf{p}). \quad (3)$$

We observe that the chromaticity of α is the albedo chromaticity, i.e.,

$$\frac{\alpha^c(\mathbf{p})}{\sum_x \alpha^x(\mathbf{p})} = \frac{\rho^c(\mathbf{p})}{\sum_x \rho^x(\mathbf{p})}. \quad (4)$$

Since we have the albedo chromaticity, we can remove its contribution from the no-flash image, and use (1) and (3) to obtain:

$$\beta^c(\mathbf{p}) = \frac{I_{nf}^c(\mathbf{p})}{\frac{\alpha^c(\mathbf{p})}{\sum_x \alpha^x(\mathbf{p})}} = \left(\sum_x \rho^x(\mathbf{p}) \right) \sum_i \lambda_i(\mathbf{p}) \ell_i^c. \quad (5)$$

Finally, we can remove the term $(\sum_x \rho^x(\mathbf{p}))$ simply by taking the chromaticity of $\beta^c(\mathbf{p})$:

$$\gamma^c(\mathbf{p}) = \frac{\beta^c(\mathbf{p})}{\sum_x \beta^x(\mathbf{p})} = \frac{\sum_i \lambda_i(\mathbf{p}) \ell_i^c}{\sum_x \sum_j \lambda_j(\mathbf{p}) \ell_j^x} \quad (6)$$

Making use of the fact that $\sum_x \ell_j^x = 1$ by definition, we can rearrange the terms to express (6) as:

$$\gamma^c(\mathbf{p}) = \frac{\sum_i \lambda_i(\mathbf{p}) \ell_i^c}{\sum_j \lambda_j(\mathbf{p})} = \sum_i z_i(\mathbf{p}) \ell_i^c, \quad (7)$$

where we denote

$$z_i(\mathbf{p}) = \frac{\lambda_i(\mathbf{p})}{\sum_{j=1}^N \lambda_j(\mathbf{p})} \quad (8)$$

as the relative shading term. This term captures the fraction of the shading at a scene pixel that comes from one light source, relative to all the light sources, hence the term *relative shading*. Figure 2(b) shows the extracted Γ image for one particular no-flash/flash image pair. Note that, $z_i(\mathbf{p}) \geq 0$ since $\lambda_k(\mathbf{p})$, the shading associated with each light source at a pixel, are non-negative. Also note that, $\sum_i z_i(\mathbf{p}) = 1$ by construction.

We can now formalize the following statement that is central to the techniques proposed in this paper.

PROPOSITION 1 (THE HULL CONSTRAINT). *The term $\Gamma(\mathbf{p}) = [\gamma^r(\mathbf{p}), \gamma^g(\mathbf{p}), \gamma^b(\mathbf{p})] \in \mathbb{R}^3$, given as,*

$$\Gamma(\mathbf{p}) = \sum_{i=1}^N z_i(\mathbf{p}) \ell_i, \quad (9)$$

lies in the convex hull formed by the light source chromaticities $\{\ell_1, \dots, \ell_N\}$.

The hull constraint is simply a consequence of $z_i(\mathbf{p})$ being non-negative and $\sum_i z_i(\mathbf{p}) = 1$. Note that the value of $\Gamma(\mathbf{p})$ can be computed from a flash/no-flash image pair and the color of the flash lighting. Further, even though $\Gamma(\mathbf{p})$ is a 3-vector, it belongs to the 2D simplex since:

$$\sum_c \gamma^c(\mathbf{p}) = \sum_c \sum_i z_i(\mathbf{p}) \ell_i^c = \sum_i z_i(\mathbf{p}) \sum_c \ell_i^c = 1$$

To recap, we have derived a per-pixel property, $\Gamma(\mathbf{p})$, that can be computed from a no-flash/flash image pair with knowledge of the flash color (see Fig. 2(b) for an example Γ visualized as an image). $\Gamma(\mathbf{p})$ is a convex combination of the global scene illuminant chromaticities, where the weights of this convex combination are the per-pixel relative shadings of the scene illuminants. In the next section, we show that both the global light source chromaticities

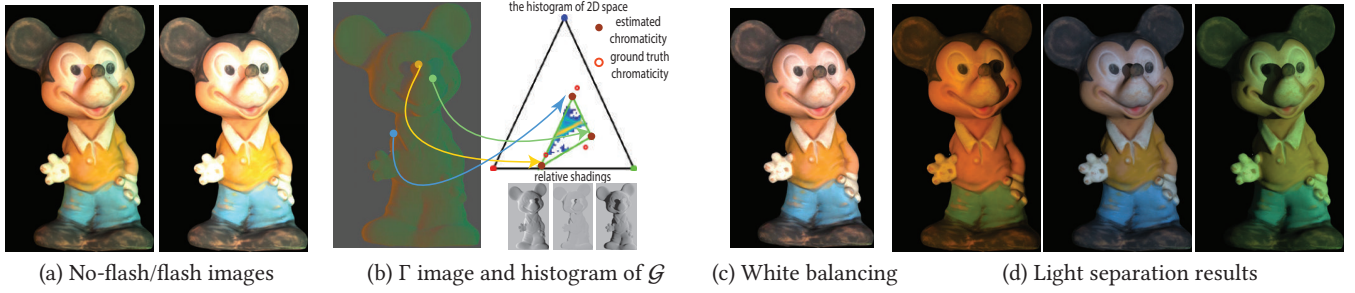


Fig. 2. Visualization of our analysis pipeline. (a) Input images. (b) Γ image as well as the \mathcal{G} , the histogram/scatter plot of $\{\Gamma(\mathbf{p}), \forall \mathbf{p}\}$ on the 2D simplex of chromaticities. Notice how the scatter plot lies within the convex hull of the light source chromaticities. Also “pure” pixels in the scene that are lit by only single light sources project to the corners of this convex hull. Our method utilizes the hull constraint to (b) estimate the lighting chromaticities as well as the relative shading at each pixel, (c) white balance the no-flash image, and (d) perform source separation based on lighting chromaticity.

and the per-pixel relative shading can be estimated by analyzing the set of all $\Gamma(\mathbf{p})$ values in the image, i.e., $\mathcal{G} = \{\Gamma(\mathbf{p}); \forall \mathbf{p}\}$.

4 LIGHTING ANALYSIS WITH THE HULL CONSTRAINT

Given the no-flash/flash images, I_{nf} and I_f respectively, and the flash chromaticity ℓ_f , our goal is render novel images by altering the color and intensity of the light sources in the scene. This requires us to be able to separate the no-flash image into components corresponding to each distinct light chromaticity. In this section, we will show that this is possible using the Γ image defined in Sec. 3.2. We begin by describing how we can estimate a specific type of re-rendered image – a white-balanced image.

4.1 A novel white balancing technique

Images captured under mixed illumination have casts caused by the colors of the scene lights. White-balancing seeks to remove these color casts by normalizing each scene illuminant to be white. In our context this means rendering the white-balanced image, I_{wb} , as:

$$I_{wb}^c(\mathbf{p}) = \rho^c(\mathbf{p}) \sum_{i=1}^N \lambda_i(\mathbf{p}). \quad (10)$$

In essence, the white balanced image has the same shading as observed in the no-flash image, but with the light source chromaticities set to pure white color.

Following prior work [25, 26], we aim to accomplish this by computing a per-pixel white balancing kernel $W(\mathbf{p}) \in \mathbb{R}^3$ such that:

$$W^c(\mathbf{p}) I_{nf}^c(\mathbf{p}) = I_{wb}^c(\mathbf{p}). \quad (11)$$

By combining this with (1), (7) and (10), we obtain the following closed form expression for the white balance kernel:

$$W^c(\mathbf{p}) = \frac{\sum_{i=1}^N \lambda_i(\mathbf{p})}{\sum_{i=1}^N \lambda_i(\mathbf{p}) \ell_i^c} = \frac{1}{\gamma^c(\mathbf{p})}, \quad (12)$$

i.e., the white-balance kernel is the inverse of the previously computed property, Γ .

Discussion. Note that, beyond the Lambertian assumption, this white balance technique does not impose any restrictions on the number or the nature of scene illuminants. This is in contrast to almost all previous white-balance techniques that can only handle

one or two scene illuminants. Further, the technique operates on a per-pixel basis and hence, is fast to compute and automatically handles spatially-varying lighting with no user input.

Figure 3 compares the results of three different flash-based white balancing methods on an image created using the MIT Intrinsic Images dataset [22]. We observe that our proposed technique estimates white balance kernels that closely approximate the ground truth. In particular, a key improvement over the previous techniques (including Hui et al. [26]) is that the proposed technique does not require the so called ‘intensity preservation’ constraint [11, 25, 26] that constrains the white balanced image to have the same luminance as the no-flash image. This constraint does not have a physical basis and leads albedos erroneously bleeding into the white-balance kernel as can be seen in the result from Hui et al. [26] (Fig. 3(e)).

4.2 Identifying light source chromaticities

Recall, from Proposition 1, that $\Gamma(\mathbf{p})$ lies in the convex hull formed by the lighting chromaticities $\{\ell_1, \dots, \ell_N\}$. This is illustrated in Figure 2. We now describe methods to estimate the number of light sources and their respective chromaticities from the set $\mathcal{G} = \{\Gamma(\mathbf{p}); \forall \mathbf{p}\}$. Intuitively, our methods rely on fitting the tightest convex hull to the set \mathcal{G} and identifying the corners of the estimated convex hull as the lighting chromaticities. Along the way, we state sufficient/necessary conditions when the resulting estimates are meaningful. We begin by discussing the identifiability of a light source.

Identifiability of a light source. A light source is identifiable only if its chromaticity lies outside the convex hull of the remaining light source chromaticities, i.e., if its chromaticity cannot be explained as a convex combination of the remaining light chromaticities. If this were not that case, then its contribution to a scene point can be explained by the remaining lights. Hence, only light sources whose chromaticities lie at corners of the convex hull of $\{\ell_1, \dots, \ell_N\}$ are identifiable given the flash/no-flash image pair. Without any loss in generality, we assume that all light sources are identifiable. Therefore, if we can identify the convex hull of the light sources $\mathcal{L} = \text{convex-hull}\{\ell_1, \dots, \ell_N\}$, we can estimate the light source chromaticities as the corner points of this set.

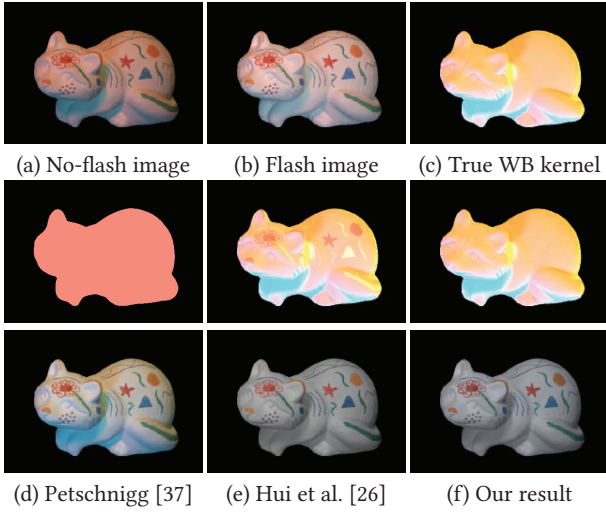


Fig. 3. We synthesize no-flash (a) (with four illuminants) and flash (b) images from dataset of Grosse et al. [22]. We compare the output white-balance kernels (middle) and final results (bottom) of three flash-based white-balance algorithms. Petschnigg et al. [37](d) assume a single illuminant and cannot handle mixed illumination. Hui et al. [26] (e) uses an intensity preservation constraint that leads to albedo bleeding into the white-balance kernel. In contrast, our result (f) closely matches the ground truth and produces results that leave the albedo of the scene unchanged.

While we do not have an a-priori estimate of \mathcal{L} , we can compute the Γ image and construct the set of all values $\mathcal{G} = \{\Gamma(\mathbf{p}); \forall \mathbf{p}\}$. Recall that, under ideal imaging conditions, Proposition 1 (Eqn. 9) states the Γ values are convex combinations of the light source chromaticities, $\{\ell_1, \dots, \ell_N\}$. This means that the Γ values are contained within the convex hull \mathcal{L} , or equivalently $\mathcal{G} \subseteq \mathcal{L}$. We next explore sufficient conditions under which the convex hull of \mathcal{G} is equal to \mathcal{L} . When these conditions are satisfied, we can estimate the light source chromaticities as the corner points of the convex hull of \mathcal{G} .

PROPOSITION 2 (PRESENCE OF “PURE” PIXELS). *Under ideal imaging conditions (absence of noise, non-Lambertian surfaces, etc.), the convex hull of \mathcal{G} is equal to \mathcal{L} if, for each light source, there exists a pixel that is purely illuminated by that light source, or, equivalently,*

$$\forall \ell \in \{\ell_1, \dots, \ell_N\}, \exists \Gamma(\mathbf{p}') = \ell.$$

When there are pure pixels for each light source, then the set \mathcal{G} will include the lighting chromaticities which are also the corners of the convex hull \mathcal{L} . Therefore, the convex hull of \mathcal{G} will be identical to \mathcal{L} . Pure pixels can be found in shadow regions since shadows indicate the absence of light source(s). The pure pixel assumption is thus satisfied when the scene geometries are sufficiently complex to exhibit a wide array of cast and attached shadows. The more complex the scene geometry, the more likely it is that we satisfy the condition in Proposition 2.

Under ideal conditions, we can identify the lighting chromaticities by detecting pure-pixels as points in \mathcal{G} that cannot be expressed

as convex combinations of the remaining points. However, this approach is extremely sensitive to non-idealities in the imaging process including measurement noise and the presence of non-Lambertian materials. Instead, we propose techniques that are specialized to the number of light sources in the scene, and are robust to such non-idealities.

Pre-processing. To reduce the effect of measurement noise and non-Lambertian materials, we reject outliers in \mathcal{G} by looking at the histogram of points on the simplex and removing points in sparsely populated regions. Specifically, we build a histogram of \mathcal{G} by dividing the 2D simplex into 100×100 bins and counting the occurrence of $\Gamma(\mathbf{p})$ in each bin. Bins that have less than 100 pixels are removed from \mathcal{G} . An example of this histogram is shown in Figure 2.

One light scenario ($N = 1$). When the scene is illuminated by a single light source, the set \mathcal{L} reduces to a single point that coincides with the chromaticity of the light source. In the presence of noise, we can expect \mathcal{G} to be clustered around the chromaticity of the light source. Hence, we can estimate this chromaticity as the mean of \mathcal{G} .

Two-lights scenario ($N = 2$). When the scene is illuminated by two light sources, the set \mathcal{L} is the line segment between the two light chromaticities, ℓ_1 and ℓ_2 . The set \mathcal{G} is, hence, clustered around this line. We use RANSAC [21] to robustly determine the line. We project \mathcal{G} onto this line and associate the end points with the light source chromaticities. The end points correspond to the true light source chromaticities only if there were “pure pixels” in the no-flash photograph that were illuminated individually by each of the light sources.

Three-lights scenario ($N = 3$). For scenes illuminated with three lights, the problem reduces to finding a triangle that encompasses \mathcal{G} ; once we find a satisfactory triangle, its vertices correspond to the light source chromaticities. This problem, by itself, is ill-constrained since there can be an infinite number of triangles that encompass \mathcal{G} and further, the simplex itself is a trivial solution for any \mathcal{G} . To resolve this ambiguity, we constrain the problem by determining the triangle with the minimum area that encompasses \mathcal{G} . This problem has been studied in great detail in the field of computational geometry and there are many solutions in current literature [3, 20, 29, 33, 35]. We leverage the method proposed by Parvu et al. [35] to determine the triangle and the associated vertices. This approach works by estimating the convex hull to \mathcal{G} and subsequently, finding three sides of the convex hull that can provide the tightest enclosing triangle. In practice, this technique works well when we have sufficient samples along the edges of \mathcal{L} — a condition that is satisfied when the scene contains pixel illuminated by all pairs of light sources, i.e., are in the shadow of one particular light.

More than three light sources ($N \geq 4$). While the procedure used for three light sources can potentially be applied to any arbitrarily higher number of sources, we can expect the method to get increasingly brittle as N increases. Further, as we will see next, even if we can robustly estimate the light source chromaticities, lighting separation can be done uniquely only when $N \leq 3$. Hence, we limit the applicability of our techniques to the scenario when $N \leq 3$. In practice, we have found that it is extremely rare for a scene to have more than three light sources with distinct chromaticities. The vast

majority of real-world scenes have two illuminants, and a small subset have three.

Identifying the number of light sources. The analysis above also provides an elegant way to identify the number of light sources N . We can estimate the mean and covariance matrix of the set \mathcal{G} . If the singular values of the covariance matrix are small, then \mathcal{G} is tightly clustered around its mean and hence, $N = 1$. If the covariance matrix is well approximated by a rank-1 matrix, then \mathcal{G} lies close to a line and hence, $N = 2$. If the covariance matrix is rank-2, with both singular values sufficiently large, then $N \geq 3$.

Once we have an estimate of the convex hull \mathcal{L} , we can associate its corners with the light source chromaticities. We denote the estimates as $\{\hat{\ell}_1, \dots, \hat{\ell}_N\}$.

4.3 Lighting separation

Given the Γ image and the estimated light chromaticities, we now describe how we explicitly separate the no-flash image into components corresponding to the individual light sources.

Estimating the relative shading. We start by estimating the relative shading corresponding to each light source. Recall that the relative shading at a pixel, $z_i(\mathbf{p})$, $i = 1, 2, \dots, N$, defined in (8), measures how much each light source contributes to the pixel color relative to all the other light sources. We can estimate it by solving for the convex combinations that produces $\Gamma(\mathbf{p})$ from $\{\hat{\ell}_1, \dots, \hat{\ell}_N\}$. Specifically, we solve the following optimization problem:

$$\Gamma(\mathbf{p}) = \sum_{i=1}^N z_i(\mathbf{p}) \hat{\ell}_i, \quad \sum_{i=1}^N z_i(\mathbf{p}) = 1.$$

We denote the relative shading estimates as $\hat{z}_i(\mathbf{p})$. Given that $\Gamma(\mathbf{p}) \in \mathbb{R}^3$, the above problem has a unique closed-form solution only when $N \leq 3$. This is a limitation that stems from using three-color cameras; in general, if the camera were to have C independent color channels, then we can uniquely obtain the relative shading terms for up to C light sources. Fig. 2(b) shows the relative shading estimated for one particular scene.

Lighting separation. Once we have the lighting chromaticities and the relative shading, we can separate the no-flash photograph into N photographs. Specifically, the separated image corresponding to the estimated chromaticity $\hat{\ell}_j$ is given as:

$$\hat{I}_{sep,k}^c(\mathbf{p}) = \rho^c(\mathbf{p}) \lambda_k(\mathbf{p}) \hat{\ell}_i^c. \quad (13)$$

Combining with (8) and (10), the separated image is rendered as:

$$\hat{I}_{sep,k}^c(\mathbf{p}) = \hat{z}_k(\mathbf{p}) I_{wb}(\mathbf{p}) \hat{\ell}_i^c. \quad (14)$$

Thus the separated image can be obtained by simply multiplying the white-balance image (which already counts for the sum total of light intensity arriving at a pixel) by the relative contribution of individual light sources, and coloring it with the estimated chromaticity. This is illustrated in Figure 2(d).

4.4 Computational requirements

The computational requirements of our approach are light since most of the steps involve per-pixel operations. The sole exception is the lighting estimation step where all the pixel chromaticities are analyzed. For the two-lights scenario, our unoptimized MATLAB

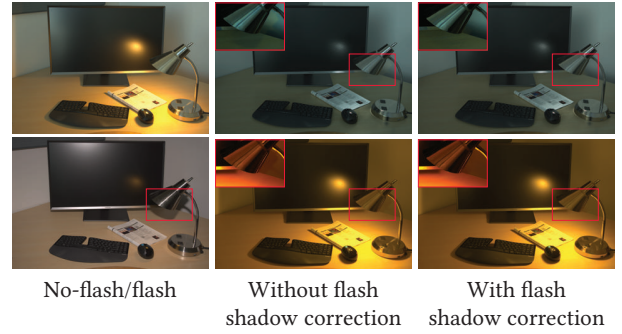


Fig. 4. We show flash shadows correction by comparing with the lighting separation results. In the scene, the desk lamp casts shadows on the wall, leading to the artifacts in the lighting separation results as shown in the close-up rendition. By correcting the artifacts, our method produces high quality results.

implementation takes ≈ 1 minute to separate a 1200×1900 image on a desktop with Intel Xeon 3.6G CPU. Of this, 30 seconds are spent on the RANSAC line-finding. For three-lights scenario, the algorithm to find the minimum bounding triangle converges much more faster, for a total of ≈ 30 seconds.

4.5 Handling flash shadows

Our method assumes that the flash illuminates all the pixels in the scene. However, this assumption is violated in the flash shadows regions which leads to artifacts in the final results. To correct these, we first detect flash shadows using the method proposed in Hui et al. [26]. The estimates of Γ in these regions are erroneous and we correct them by propagating values from pixels outside these detected regions that are spatially close and have similar no-flash intensities. Once this is done, the rest of the estimation remains the same as described before. Figure 4 illustrates the effect of flash shadows and our corrected result.

5 APPLICATIONS AND RESULTS

In this section, we report the performance of our technique on a wide-range of real-world scenes for a number of different analysis and editing tasks. We also compare our results with task-specific state-of-the-art methods. The accompanying supplementary material contains more results and comparisons. We encourage readers to zoom into the figures to appreciate the details of the results.

We focus on two sets of applications; the first on illumination chromaticity manipulation in the form of white balancing and light color editing, and the second on intrinsic image estimation and photometric stereo. The motivation for the latter set stems from the following interesting observation. When $N = 3$, we can obtain four distinctly illuminated photographs (3 separated photographs and the pure flash image) of the scene just from two photographs (the flash/no-flash pair) by using our source separation technique. The availability of four photographs, from the flash/no-flash pair, leads to dramatic improvements in robustness in the context of intrinsic image estimation and photometric stereo.

5.1 White balance

The vast majority of white-balance algorithms assume that the scene is lit by a single dominant light source. In contrast, we leveraging an additional flash photography, we are able to estimate and remove the effect of spatially-varying lighting. Our results are shown in Figures 3 and 5, with more results in the supplementary material. We compare our results with those from two algorithms that are designed to handle spatially-varying mixed illumination — Hsu et al. [25] and Hui et al. [26]. Hsu et al. require that the color of the illuminants and assume that only two light sources present in the scene. While we manually specified this as input to their technique, their result is not able to deal with extreme illumination (see Fig. 5). Similar to us, Hui et al. use a flash camera and can generalize to an arbitrary number of scene illuminants. However, as mentioned in Sec. 4.1, they still rely on the image intensity preservation constraint [11, 25], which leads to errors in the white balance kernels.

5.2 Light source separation

We evaluate the performance of our light source separation on a number of real-world captured images.

Scenes with two lights. In Figure 6, we demonstrate our technique on a scene with two lights sources and compare it with ground truth captures. Our technique is able to successively estimate the two illuminants in the scene – warm indoor lighting and cool outdoor lighting. In addition, our technique is also to separate the no-flash image into its individual components and accurately captures all the complex lighting effects in this image (including complex shadows, intricate geometry, varied lighting).

In Figure 6, we compare our technique to two different baseline algorithms. This scene is lit by warm conference room lights and also cool outdoor illumination. Ground truth photographs were obtained by turning off the indoor light sources to obtain the outdoor illuminated scene and then subtracting this from the no-flash image to obtain the photograph with respect to the indoor illumination. We compare against a simple non-negative matrix factorization as well as the technique proposed in Hsu et al. [25]. Our technique produces the results closely resemble to the actual captured photographs, indicating its robustness and effectiveness.

We also evaluates the performance of our technique against the related light source separation methods in Figure 7. Naively applying Non-Negative Matrix factorization (NNMF) to the no-flash image leads to the loss of the colors. Hsu et al. [25] focused on estimating relative contribution of the light source from no-flash image only at the expense of introducing restrictive assumptions on the scene as well as the colors of the illuminants. In the paper, we manually set the light colors for the best performance. As can be seen in Figure 7, the method of Hsu et al. leads to the visual artifacts by incorporating the strong prior on the scene. In contrast, our approach requires very few assumptions on the underlying image and returns the results most close to the ground truth. Comparisons with Hsu et al. [25] on more real and synthetic scenes can be found in supplementary material.

Sunlight and skylight separation. An interesting application of two-light source separation is in outdoor time lapse videos where it

is often necessary to separate direct sunlight from indirect skylight. Figure 8 showcases the performance of light separation technique on an outdoor scene. For this scene, we identify a photograph with cloudy sky where there is no direct sunlight and the entire scene is lit only by the skylight, which is an area source and illuminates the scene uniformly. We use this as a pure flash photograph; it is worth emphasizing that our technique does not make any assumptions about the nature of the flash illumination; here, we exploit it to use skylight in place of the flash illumination. Also note that skylight changes its color and intensity significantly during the course of the day. Given this pure flash photograph, our separation scheme is able to produce the results closely resemble to the manner of the sky and the sun illumination.

Scenes with three lights. Figure 9 showcase our techniques on scenes with three lights. The scene in Fig 9 (top-row) is illuminated by the outdoor illuminant but the curtain on the window introduces a second light source that is tinged with red. In addition to these two sources, the ceiling light is also turned on in the room. While the scene features complex shadows (under the chair, around the small table, near the red curtain, to name a few) caused by different type of illuminants, our lighting separation scheme is able to group these regions into the associated separated photographs and produce visually pleasing results. In Figure 9 (bottom-row), the building blocks are illuminated by the light sources with three light colors, i.e. an LED light with light green, a fluorescent lamp with yellow colors and sky light from the windows. The consistency of shadows in the images showcases the robustness and success of our light separation technique.

Light color and brightness editing. Given the separated results, we can adjust the brightness as well as the color chromaticity of particular light simply by operating on $\tilde{I}_{\text{sep},k}$. Specifically, we produce the photograph under the novel illumination condition by

$$\tilde{I} = \sum_k \tilde{I}_{\text{sep},k} \alpha_k \tilde{\ell}_k, \quad (15)$$

where $\tilde{\ell}_k$ denotes the adjusted chromaticity and α_k denotes the changes in the brightness. Figure 10 shows two examples of editing the light color and brightness for the captured ambient image. We experiment by adjusting the parameters α and $\tilde{\ell}_k$ in (15). The rendered photographs are both visually pleasing and photorealistic in their preservation of shading and shadows.

5.3 Intrinsic images

Intrinsic image algorithms seek to separate an image into a per-pixel product of reflectance, that represents surface color, and shading, which captures the effect of scene illumination and geometry. This problem is highly ill-posed since two unknowns are required to be estimated from a single observation. A common approach to restrict the solution space is to incorporate the priors on both reflectance and illumination [5, 7, 48]. Most single-image intrinsic decomposition methods assume that scene is illuminated by a single white light source and that the shading image is spatially smooth. Both these assumptions fail on real-world scenes that can have spatially-varying mixtures of illuminants, and high-frequency illumination that lead to shadows and other discontinuities in shading.

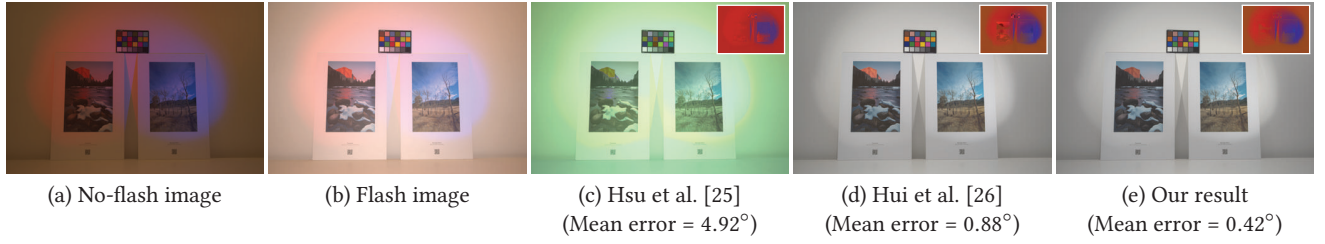


Fig. 5. We evaluate our white balance method on a no-flash/flash pair (a/b) from Hui et al. [26]. We show the white balance results and the associated kernels (insets). (c) Hsu et al. [25] require the light colors to be manually specified but fail on the extreme illumination in this scene. (d) Hui et al. use a flash image to improve results, but their use of ‘intensity preservation’ leads to albedos bleeding into the white balance kernel (zoom into the insets). (e) Our method produces the best result in terms of both visual quality and angular error.



Fig. 6. We evaluate our lighting separation results by comparing against actual captured photographs. Specifically, we take photographs under the outdoor illuminant only by turning off all the indoor light sources. As can be seen here, our method produces results that are very close to the captured photographs.

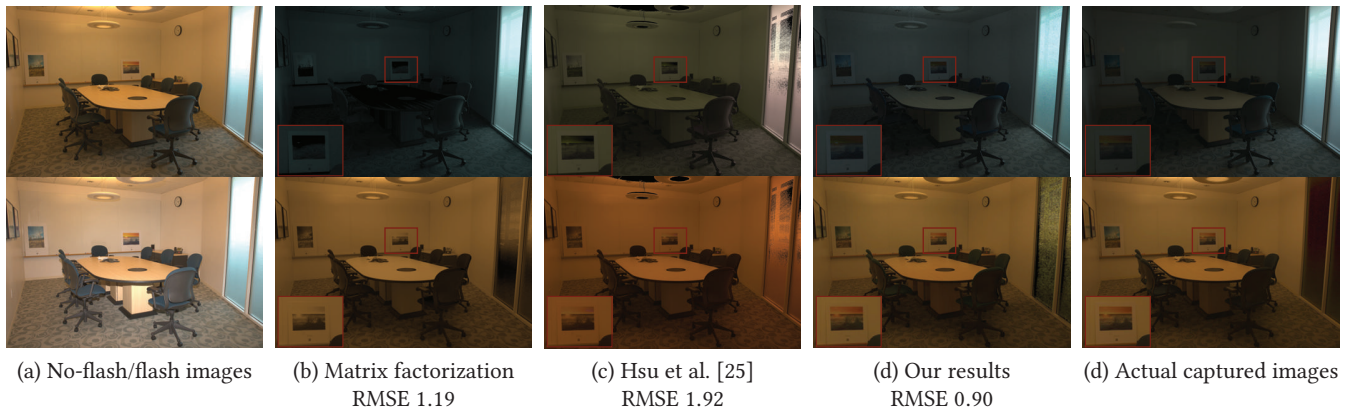


Fig. 7. We separate a no-flash image (a) into two components and compare with matrix-factorization (b) and Hsu et al. [25] (c). Compared to the ground truth images, we can see that matrix factorization produces noisy colors (see the painting on the right), while Hsu et al. [25] produce an incorrect estimate of light color and shading. Our result (d) closely mimics the actual captured results.

We can directly use our estimates of spatially-varying light chromaticity to account for variations in illumination color. In addition, we are already separating the no-flash image into multiple images,

each of which captures different lighting. We show that we can leverage these separation results to propose a more effective shading prior that leads to state-of-the-art intrinsic decompositions.

Algorithm. Following the standard intrinsic image formulation, the no-flash image can be denoted as $I_{nf}(\mathbf{p}) = \mathbf{R}(\mathbf{p})\mathbf{S}_{nf}(\mathbf{p})$, where

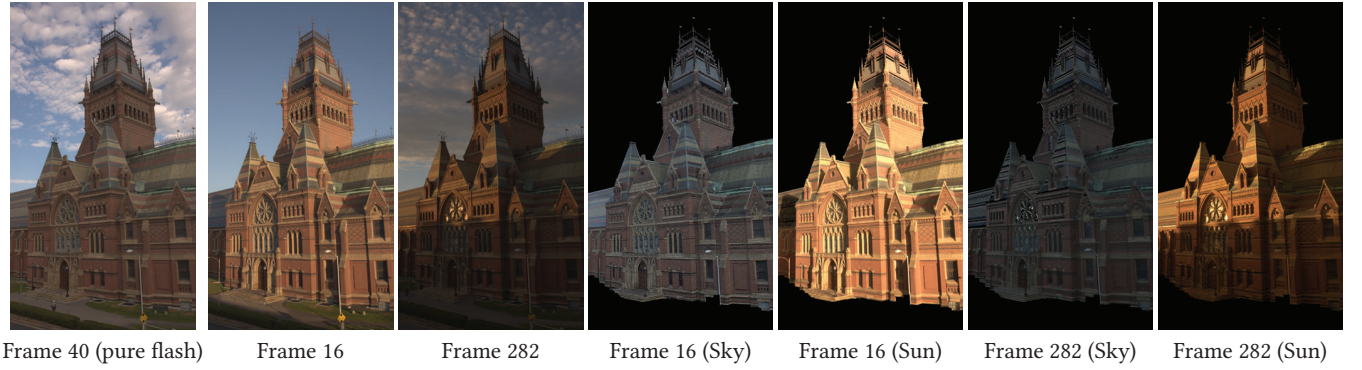


Fig. 8. We demonstrate our technique on the time lapse sequence under the mixture of the Sun and Sky illuminations. We use the cloudy day frame as the pure flash image. Note that the Sun can be characterized as the directional light source, which introduces the cast shadows in the scene, while the sky can be considered as the area light. As can be seen from the separated images, our algorithm is able to produce good results, especially in encoding the shading attributes for both light sources in the scene.



Fig. 9. We evaluate our technique on scenes with mixtures of three lights. In the top row, we capture an image under warm indoor LED lights and outdoor lighting that percolates through a red curtain. The outdoor light that is transmitted through the curtain gets colored red, while some of it bounces off the curtain's white backing and diffuses into the scene as blue light. Our technique is able to estimate separated results that capture this complex light transport. In the bottom image, we image a scene under warm indoor lighting, a green fluorescent lamp and cool skylight. Our separation results capture both the color and the shading for each of these sources.

$\mathbf{R}(\mathbf{p})$ and $\mathbf{S}(\mathbf{p})$ denote the scene reflectance and no-flash shading, both of which are 3-color vectors in our case. By taking log, this reduces to the linear form: $\mathbf{i}_{nf}(\mathbf{p}) = \mathbf{r}(\mathbf{p}) + \mathbf{s}(\mathbf{p})$. Intrinsic image techniques solve for $\mathbf{r}(\mathbf{p})$ and $\mathbf{s}(\mathbf{p})$ by minimizing:

$$E(\mathbf{p}) = \sum_{\mathbf{p}} \|\mathbf{i}_{nf}(\mathbf{p}) - \mathbf{r}(\mathbf{p}) - \mathbf{s}(\mathbf{p})\|^2 + E_s(\mathbf{p}) + E_r(\mathbf{p}), \quad (16)$$

where the first term is the reconstruction error and the second and third terms are prior on the shading and reflectance. We now describe our novel shading-prior.

Shading prior. Each of our separated images captures the scene as lit by a different illuminant. If two pixels in the scene are lit in the same way in each of these images, it would indicate that they have similar geometric properties like the same surface normal and the same occlusion map, and therefore are likely to have similar shading values. This property has been used before in Photometric



Fig. 10. We separate no-flash images (a) into individual light components, and recolor them to create photo-realistic results with novel lighting conditions (b). Note how our method changes the color and brightness of each light while realistically retaining all shading effects.

Stereo methods [24] to recover surface normals. We propose using this to build a shading prior for intrinsic images.

Given the lighting separation results, we denote the intensity profile at \mathbf{p} as $v(\mathbf{p}) = [\hat{i}_{\text{sep},1}(\mathbf{p}) \cdots \hat{i}_{\text{sep},N}(\mathbf{p})]^T$, where \hat{i} denotes the grayscale image of the corresponding separated images. Next, we normalize $v(\mathbf{p})$ by dividing the $\|v(\mathbf{p})\|_2$ across the separated Q image to remove the effect of per-pixel albedo, and denote this vector at \mathbf{p} as $\mathbf{V}(\mathbf{p})$. We call $\mathbf{V}(\mathbf{p})$ the shading guide. We enforce neighboring pixels that have similar values of $\mathbf{V}(\mathbf{p})$ to have a similar shading value by using the following shading prior:

$$E_s(\mathbf{p}) = \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \exp\left(\frac{-\|\mathbf{V}(\mathbf{p}) - \mathbf{V}(\mathbf{q})\|^2}{2\sigma_s^2}\right) (s(\mathbf{p}) - s(\mathbf{q}))^2 \quad (17)$$

We are using the shading guide $\mathbf{V}(\mathbf{p})$ to evaluate how similar we expect the shading at neighboring pixels to be and scaling the local shading smoothness by it. As observed from Figure 11, the shading guide $\mathbf{V}(\mathbf{p})$ closely matches the geometry of the object, and consequently leads to an accurate shading prior.

Reflectance prior. Similar to previous work, we use a chromaticity-based prior on the reflectance. More specifically, we utilize the chromaticity from each separated image $[\hat{i}_{\text{sep},1}(\mathbf{p}) \cdots \hat{i}_{\text{sep},N}(\mathbf{p})]$ and denote them as $[C_1(\mathbf{p}) \cdots C_Q(\mathbf{p})]$. We average the N chromaticities to compute a reflectance guide image, $\mathbf{U}(\mathbf{p})$ (see Figure 11). We use this reflectance guide to compute a local reflectance prior that forces neighboring pixels with similar $\mathbf{U}(\mathbf{p})$ to have similar reflectances. In addition, we follow the technique of Boneel et al. [9] to construct a non-local prior that clusters pixels based on $\mathbf{U}(\mathbf{p})$ and forces pixels in the same clusters to have similar reflectances. We refer the reader to their paper for more details.

We use our shading and reflectance priors and solve a sparse linear system to minimize Equation 16. We weight the shading prior by 0.1 and the reflectance prior by 1.

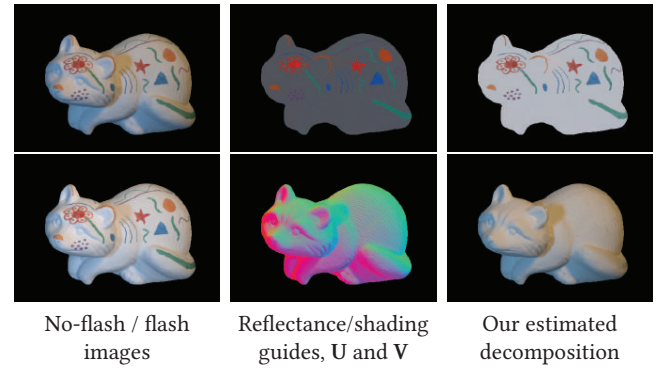


Fig. 11. We separate a no-flash image created from [22] (a) into two images and visualize our reflectance and shading guides (b, top/bottom). As can be seen, these guides closely match the true reflectance and geometry of the scene leading to high-quality reflectance and shading estimates (c, top/bottom).

Evaluation. To quantitatively evaluate our technique, we compare its performance to the state-of-the-art methods on the benchmarking dataset from Grosse et al. [22]. This dataset has images of each object under varying lighting conditions. We generated a no-flash image with a mixture of two lights by modulating two images with the pre-selected light colors, and adding them together. We also chose the image lit by a frontal light source as the pure flash image, and added it to the no-flash result to generate the flash input image. We use the no-flash/flash pair to compute our decomposition, and also run other state-of-the-art techniques on the no-flash image. As can be seen from Figure 12, we are able to produce results that are, in terms of LMSE error, an order of magnitude better than the state-of-the-art methods.

Figure 13 visually illustrates the performance of our technique for the intrinsic image decomposition on two real-world scenes and

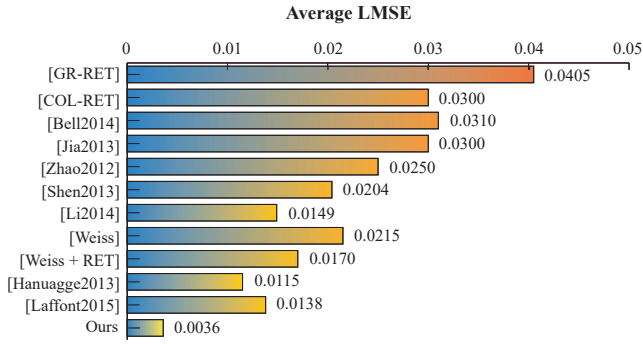


Fig. 12. We evaluate against the state-of-the-art methods in terms of Local Mean Squared Error (LMSE) on the MIT intrinsic dataset.

compares with state-of-the-art techniques. We refer readers to the supplementary material for more such comparisons. In particular, we evaluate against the single-image techniques of Shen et al. [39] and Bell et al. [7] and the multi-image technique of Weiss [44] + Retinex (Weiss + RET), and Hauagge et al. [23]. We use the no-flash image as the input to test the performance of single-image techniques, and the no-flash/flash image pair for the multi-image techniques. As can be seen in Figure 13, our technique significantly outperforms the state-of-the-art methods in the estimates for the reflectance and shading. Most of these techniques assume a single light color or one ambient white illuminant, which leads to color variations in the shading to be incorporated into the reflectance estimates. In addition, most previous methods struggle in regions with high-frequency texture; these variations are usually transferred into the shading. Unlike these approaches, our technique handles spatially-varying color illumination, and correctly separates the detail in the shading and reflectance. While our technique requires one additional image, the only intrinsic image techniques, that we are aware of, that can handle these situations either require a larger set of images [28] or significant user interaction [10]. In contrast, our results are completely automatic.

5.4 Photometric stereo

Photometric stereo [45] methods aim to surface shape (usually normals) of an object from images obtained from a static camera under varying lighting. For Lambertian objects, this requires a minimum of three images. Recently, techniques have been proposed to do this from a single shot where the object is lit by three monochromatic red, green, and blue, directional light sources [13, 14]. However this estimation is still ill-posed and requires additional priors. We propose augmenting this setup by capturing an additional image lit by a flash collocated with the camera. We use our proposed technique for lighting separation to create three images out of this (plus the flash image), at which point we can use standard calibrated Lambertian Photometric Stereo to estimate surface normals. As shown in Figure 14 this leads to results that are orders of magnitude more accurate than the state-of-the-art technique [14]. More comparisons can be seen in the supplementary material.

6 CONCLUSION

In this paper, we have shown that capturing an additional image of a scene under flash illumination, allows us to separate the no-flash image into image corresponding to illuminants with unique colors. Our technique is based on a novel colorimetric analysis that derives the conditions under which this is possible. We have also shown that this ability to analyze and isolate lights in turn leads to state-of-the-art results on tasks such as white balancing, illumination editing, intrinsic image decomposition and color photometric stereo. We believe that this is a significant step towards true post-capture lighting control over images.

Limitations and future work Our technique assumes that the scene is Lambertian and we would like to extend it to handle complex surface reflectance. In addition, while it separates the contribution of each light component, it does not do a scene reconstruction. As a result, while we can edit the color of each light source, we cannot move them in space. Enabling this form of editing would enable complete lighting control over scenes. We are also interest in exploring how algorithm can be extended to other scenarios like video, and hyper-spectral capture.

REFERENCES

- [1] Supreeth Achar and Srinivasa G Narasimhan. 2014. Multi focus structured light for recovering scene shape and global illumination. In *ECCV*. 205–219.
- [2] Amit Agrawal, Ramesh Raskar, Shree K Nayar, and Yuanzhen Li. 2005. Removing photography artifacts using gradient projection and flash-exposure sampling. In *TOG*, Vol. 24. 828–835.
- [3] Esther M. Arkin, Y-J Chiang, Martin Held, Joseph S. B. Mitchell, Vera Sacristan, SS Skiena, and T-C Yang. 1998. On minimum-area hulls. *Algorithmica* 21, 1 (1998), 119–136.
- [4] Jonathan T Barron and Jitendra Malik. 2012. Color constancy, intrinsic images, and shape estimation. In *Computer Vision—ECCV 2012*. 57–70.
- [5] Jonathan T Barron and Jitendra Malik. 2015. Shape, illumination, and reflectance from shading. *PAMI* 37, 8 (2015), 1670–1687.
- [6] H. Barrow and J. Tenenbaum. 1978. Recovering intrinsic scene characteristics from images. *Computer Vision Systems* (1978).
- [7] Sean Bell, Kavita Bala, and Noah Snavely. 2014. Intrinsic images in the wild. *TOG* 33, 4 (2014), 159.
- [8] Samuel Boivin and Andre Gagalowicz. 2001. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *SIGGRAPH*.
- [9] Nicolas Bonneel, Kalyan Sunkavalli, James Tompkin, Deqing Sun, Sylvain Paris, and Hanspeter Pfister. 2014. Interactive intrinsic video editing. *TOG* 33, 6 (2014), 197.
- [10] Adrien Bousseau, Sylvain Paris, and Frédo Durand. 2009. User-assisted intrinsic images. In *TOG*, Vol. 28. 130.
- [11] Iyaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Frédo Durand. 2012. User-guided white balance for mixed lighting conditions. *TOG* 31, 6 (2012), 200.
- [12] Iyaylo Boyadzhiev, Sylvain Paris, and Kavita Bala. 2013. User-assisted image compositing for photographic lighting. *TOG* 32, 4 (2013), 36–1.
- [13] Gabriel J Brostow, Carlos Hernández, George Vogiatzis, Bjorn Stenger, and Roberto Cipolla. 2011. Video normals from colored lights. *PAMI* 33, 10 (2011), 2104–2114.
- [14] Ayan Chakrabarti and Kalyan Sunkavalli. 2016. Single-image RGB Photometric Stereo With Spatially-varying Albedo. In *IEEE International Conference on 3D Vision*.
- [15] Paul Debevec. 2005. Making* The Parthenon. In *6th international symposium on virtual reality, archaeology, and cultural heritage*, Vol. 4.
- [16] Paul Debevec. 2008. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH 2008 classes*. 32.
- [17] Paul Debevec. 2012. The Light Stages and Their Applications to Photoreal Digital Actors. In *SIGGRAPH Asia*.
- [18] Per Einarsson, Charles-Felix Chabert, Andrew Jones, Wan-Chun Ma, Bruce Lomond, Tim Hawkins, Mark Bolas, Sebastian Sylwan, and Paul Debevec. 2006. Relighting Human Locomotion with Flowed Reflectance Fields. In *Eurographics Conference on Rendering Techniques*. 183–194.
- [19] Elmar Eisemann and Frédo Durand. 2004. Flash photography enhancement via intrinsic relighting. In *TOG*, Vol. 23. 673–678.



Fig. 13. We compare our estimated reflectance (row 1,3) and shading (row 2,4) with state-of-the-art techniques. As can be seen here, our results are significantly better than other technique. We handle spatially-varying lighting, which is often baked into the reflectance for other methods (see the lighting on the right wall, and the book). We also accurately separate texture from the shading (see the texture on the carpet and the text on the book).

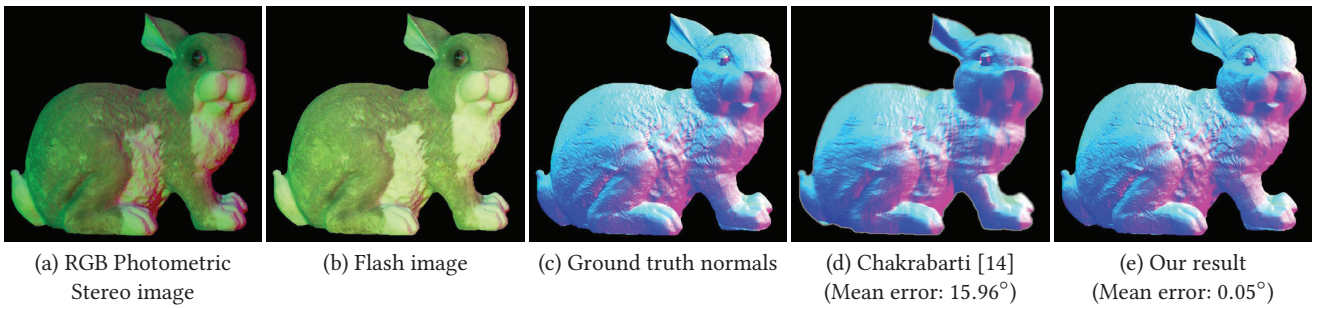


Fig. 14. Results on two-shot captured photometric stereo of real objects. We show estimated normal map for our technique as well as that of Chakrabarti et al. [14]. We also include the mean of the angular errors for the estimated surface normals. Our technique produces results that are very close to the ground truth, and significantly better than the state-of-the-art technique.

- [20] David Eppstein, Mark Overmars, Günter Rote, and Gerhard Woeginger. 1992. Finding minimum area-kgons. *Discrete & Computational Geometry* 7, 1 (1992), 45–58.
- [21] Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981), 381–395.
- [22] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. 2009. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *CVPR*.
- [23] Daniel Hauagge, Scott Wehrwein, Kavita Bala, and Noah Snavely. 2013. Photometric ambient occlusion. In *CVPR*.
- [24] Aaron Hertzmann and Steven M. Seitz. 2005. Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 8 (Aug. 2005), 1254–1264.
- [25] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Fredo Durand. 2008. Light mixture estimation for spatially varying white balance. In *TOG*, Vol. 27. 70.
- [26] Zhuo Hui, Aswin C. Sankaranarayanan, Kalyan Sunkavalli, and Sunil Hadap. 2016. White Balance under Mixed Illumination using Flash Photography. In *ICCP*.
- [27] Thomas P Koninckx and Luc Van Gool. 2006. Real-time range acquisition by adaptive structured light. *PAMI* 28, 3 (2006), 432–445.
- [28] Pierre-Yves Laffont and Jean-Charles Bazin. 2015. Intrinsic Decomposition of Image Sequences from Local Temporal Variations. In *ICCV*.
- [29] Anna Medvedeva and Asish Mukhopadhyay. 2003. An Implementation of a linear time algorithm for computing the minimum perimeter triangle enclosing a convex polygon.. In *CCCG*, Vol. 3. 25–28.
- [30] Chris Murphy, Daniel Lindquist, Ann Marie Rynning, Thomas Cecil, Sarah Leavitt, and Mark L. Chang. 2007. Low-cost stereo vision on an FPGA. In *Field-Programmable Custom Computing Machines*. 333–334.
- [31] Shree K Nayar, Gurunandan Krishnan, Michael D Grossberg, and Ramesh Raskar. 2006. Fast separation of direct and global components of a scene using high frequency illumination. In *TOG*, Vol. 25. 935–944.
- [32] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. 2005. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report* 2, 11 (2005), 1–11.
- [33] Joseph O'Rourke, Alok Aggarwal, Sanjeev Maddila, and Michael Baldwin. 1986. An optimal algorithm for finding minimal enclosing triangles. *Journal of Algorithms* 7, 2 (1986), 258–269.
- [34] Matthew O'Toole, Supreeth Achar, Srinivasa G Narasimhan, and Kiriakos N Kutulakos. 2015. Homogeneous codes for energy-efficient illumination and imaging. *TOG* 34, 4 (2015), 35.
- [35] Ovidiu Părvu and David Gilbert. 2014. Implementation of linear minimum area enclosing triangle algorithm. *Computational and Applied Mathematics* (2014), 1–16.
- [36] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec. 2007. Post-production Facial Performance Relighting Using Reflectance Transfer. *TOG* 26, 3 (2007).
- [37] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. 2004. Digital photography with flash and no-flash image pairs. *TOG* 23, 3 (2004), 664–672.
- [38] Ramesh Raskar, Kar-Han Tan, Rogerio Feris, Jingyi Yu, and Matthew Turk. 2004. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. In *TOG*, Vol. 23. 679–688.
- [39] Jianbing Shen, Xiaoshan Yang, Xuelong Li, and Yunde Jia. 2013. Intrinsic image decomposition using optimization and user scribbles. *IEEE transactions on cybernetics* 43, 2 (2013), 425–436.
- [40] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. 2006. Flash matting. *TOG* 25, 3 (2006), 772–778.
- [41] Jian Sun, Jian Sun, Sing Bing Kang, Zong-Ben Xu, Jian Sun, and Heung-Yeung Shum. 2007a. Flash Cut: Foreground Extraction with Flash and No-flash Image Pairs. *CVPR* (2007).
- [42] Jian Sun, Jian Sun, Sing Bing Kang, Zong-Ben Xu, Xiaoou Tang, and Heung-Yeung Shum. 2007b. Flash cut: Foreground extraction with flash and no-flash image pairs. In *CVPR*.
- [43] Minh Vo, Srinivasa G Narasimhan, and Yaser Sheikh. 2016. Texture Illumination Separation for Single-Shot Structured Light Reconstruction. *PAMI* 38, 2 (2016), 390–404.
- [44] Yair Weiss. 2001. Deriving intrinsic images from image sequences. In *ICCV*.
- [45] Robert J Woodham. 1980. Photometric method for determining surface orientation from multiple images. *Optical engineering* 19, 1 (1980), 191139–191139.
- [46] Ruigang Yang and Marc Pollefeys. 2003. Multi-resolution real-time stereo on commodity graphics hardware. In *CVPR*.
- [47] Yizhou Yu, Paul Debevec, Jitendra Malik, and Tim Hawkins. 1999. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *SIGGRAPH*.
- [48] Qi Zhao, Ping Tan, Qiang Dai, Li Shen, Enhua Wu, and Stephen Lin. 2012. A closed-form solution to retinex with nonlocal texture constraints. *PAMI* 34, 7 (2012), 1437–1444.
- [49] Shaojie Zhuo, Dong Guo, and Terence Sim. 2010. Robust flash deblurring. In *CVPR*.